Instructions for the Calculation of Weighted 95% UCLs for a Combined Decision Unit from Several Smaller Decision Units

This spreadsheet is a tool for use in accordance with the Interstate Technology & Regulatory Council (ITRC) document: *Incremental Sampling Methodology* (ITRC 2012).

Specifically, the calculator can be used to perform calculations in Section 4.4.1 of the ITRC document that allow the calculation of weighted means, standard errors, and 95% upper confidence limits (UCLs) for a combined decision unit (DU) made up of two or more smaller DUs.

Two types of 95% UCLs are calculated. The 95% Student's t UCL is calculated for data sets with low variability within DUs, while the 95% Chebyshev UCL is calculated for data sets with high variability within decision units (ITRC 2012). Section 4.3.4 of Incremental Sampling Methodology (ITRC 2012) discusses the performance metrics for the UCL that can be used to determine which UCL method may be more likely to achieve the study objectives. There are also considerations based on the size of the data set that affect which UCL is recommended.

The calculator recommends the 95% Student's t UCL for data sets where the coefficent of variation (CV) is low (defined as CV < 1.5), and the 95% Chebyshev UCL for data sets where the CV is medium (defined as 1.5 < CV < 3) or high (defined as CV > 3). Lognormal distrubutions may also be investigated for their suitability. If the data distribution is lognormal, the geometric standard deviation (GSD) may be used to evaluate the possibility of using the 95% Student's t UCL. The calculator does not include these calculations; see ITRC 2012 for guidance.

The method used to estimate the coefficient of variation of the individual and combined DUs is an approximation. In addition, implementation of the Welch-Satterthwaite approximation in the calculation of the 95% Student's t UCL is recommended only in cases where only moderate departures from normality are observed. The ITRC Incremental Sampling Methodology (ISM) Workgroup has not thoroughly tested the performance of this approach under conditions where distributions are highly skewed and weighting factors are very different.

Inputs

The green-shaded cells in the calculator indicate cells that the User should provide input for:

9638123

Select whether the sampling units (SU) are based on area or volume - this field should be set to either area or volume to represent what the SUs are based on (usually this will be area).
Number of increments per replicate sample - this is the number of increments present in each replicate sample. For data sets where the number of increments is not the same for all samples, it is recommended that a statistician be consulted.
DU Name - this field is optional, but provides a place for the User to add a descriptive name for each DU.
DU Area - this field is required. The User should enter the DU surface area or DU volume (in any consistent units) for each DU.

Replicate concentration - these fields are required. Although there are spaces for up to 10 DUs and 5 replicates per DU, not all of the cells must have entries. At a minimum, three replicates should be entered for each DU (as a reminder, the text for each DU will remain red until three replicates have been entered for that DU). It is not required to have the same number of replicates for each DU. At least two DUs should be entered into the calculator.

Outputs
The blue-shaded cells in the calculator show outputs for the individual DUs that the ITRC guidance recommends for evaluating combined DUs:

Number of replicates for each DU - this is the total number of replicate samples collected from each DU. This includes any sample that may be referred to as an "original" sample.

Weight for each DU - the fraction of the total combined DU that corresponds to each individual DU.
Arithmetic mean for each DU - the sum of the concentrations for all replicate samples in the DU divided by the number of samples in the DU.

SD of replicates for each DU - the standard deviation (SD) of the replicates for the specific DU.

SD of increments for each DU - the SD of all the increments within the specific DU. Because the data for the increments is not available, the SD for the DU is calculated by multiplying the SD for the replicates by the square root of the total number of increments within each replicate.
CV of DU for each DU - the CV for the specific DU.

SE for each DU - the standard error (SE) for the specific DU. The standard error is the SD divided by the square root of the number of samples for the DU.
Student's-t 95% UCL - the 95% UCL for the Student's t method. See ITRC 2012, Section 4.4.1 for the equation.
Chebyshev 95% UCL - the 95% UCL for the Chebyshev method. See ITRC 2012, Section 4.4.1 for the equation.
The purple-shaded cells show outputs for the combined DU, weighted by area.
DU Area - the total area for the combined DU.
Total number of replicates - this includes the original sample as well as all replicates.
Weight - the total weight. This should always be 1; if it is not then there is an error in the calculations.
Mean - the weighted mean for the combined DU.
SD - the weighted SD for the combined DU.
SE - the weighted SE for the combined DU.
CV - the CV for the combined DU.
Student's-t 95% UCL - the weighted 95% UCL for the Student's t method. See ITRC 2012, Section 4.4.1 for the equation.
Chebyshev 95% UCL - the weighted 95% UCL for the Chebyshev method. See ITRC 2012, Section 4.4.1 for the equation.
Degrees of freedom by Welch-Satterthwaite approximation - used to calculate the Student's-t 95% UCL. See ITRC 2012, Section 4.4.1 for the equation (ITRC 2012). Note that the Welch-Satterthwaite approximation relaxes the assumption of equal variance of the DUs, but still carries the assumption of normality.

Recommended UCL - the recommendation is based on the dispersion of the individual DUs; see Table 4-4 of ITRC 2012 for a more detailed explanation of the recommendations.

Notes
If there are non-detect results, the User can optionally enter them as the detection limit divided by two to estimate the UCL. This method does not provide results that are as accurate and precise as results from the preferred Kaplan-Meier method (EPA 2010). It is recommended that a statistician be consulted for assistance with calculating weighted 95% UCLs for data sets with non-detect results.

References

ITRC. 2012. Technical and Regulatory Guidance, Incremental Sampling Methodology. February.
EPA. 2010. "ProUCL Version 4.1.00 Technical Guide (Draft)." Prepared by Singh, A. and A.K. Singh. EPA/600/R-07/041. May. Available online at: http://www.epa.gov/osp/hstl/tsc/ProUCL_v4.1_user.pdf

Calculation of Weighted 95% UCLs for a Combined Decision Unit (DU) from Several Smaller DUs

See the "Instructions" tab (worksheet) for detailed instructions.

Project ID:
Property/Sample ID:
Date:
Operator:

	Select from drop-down / Type in number
Select whether the DU is based on area, volume or depth:	Area
Number of increments per replicate:	30

DU	DU Name	DU Area (any constant units)	Replicate concentration					Number of Replicates	Weight	Arithmetic Mean	SD of replicates	SD of increments	CV of DU	SE of DU	95% UCL	
			Rep 1	Rep 2	Rep 3	Rep 4	Rep 5								Student's-t	Chebychev
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
Combine the Weighted DUs			--	--	--	--	--						Low			

Must add to 1.00

Degrees of freedom by Welch-Satterthwaite approximation	

Student's-t or Chebychev 95% UCL may be appropriate.	Recommended UCL
	Student's t 95% UCL
The User should consult the instructions for additional guidance on which 95% UCL is recommended for specific data sets.	

For calculation purposes only

W^2	SD^2	1/n	W^2*SD^2/n

Notes

DU	Decision unit
SD	Standard deviation
SE	Standard errror
UCL	Upper confidence limit

When there are multiple samples in each strata, the overall mean of the larger DU can be estimated using the following formulae. Let n_i represent the number of samples from region i, $x-bar_i$ represent the mean of the ISM samples from region i, s_i represent the SD of the replicate ISM samples from region i, and w_i represent the weight, i.e., the relative size associated with region i. Note that if all strata are of the same size, the w_i are equal, and these equations simplify to the more common calculation methods for the mean and standard deviation. The relative size is the percentage of the larger DU that is made up of region i. The weighted mean is thus:

$$\text{Weighted Mean} = \sum_i w_i \bar{x}_i$$

The standard error associated with the weighted mean is:

$$\text{Standard Error} = \sqrt{\sum_i w_i^2 \frac{s_i^2}{n_i}}$$

which has degrees of freedom approximated by the Welch-Satterthwaite approximation (Cochran, 1977):

$$df \approx \frac{\left(\sum_i \frac{w_i^2}{n_i} s_i^2\right)^2}{\sum_i \frac{\left(\frac{w_i^2}{n_i} s_i^2\right)^2}{n_i - 1}}$$

Calculation of Weighted 95% UCLs for a Combined Decision Unit (DU) from Several Smaller DUs

See the "Instructions" tab (worksheet) for detailed instructions.

Project ID: West Co. Road; Cr(VI) project
Property/Sample ID: GW-050 (3 rep ISs per SU)
Date:
Operator:

Select from drop-down: Select whether the DU is based on area, volume or depth: **Depth**

Type in number: Number of increments per replicate: **30**

DU	DU Name	DU Depth (any constant units)	Replicate concentration					Number of Replicates	Weight	Arithmetic Mean	SD of replicates	SD of increments	CV of DU	SE of DU	95% UCL	
			Rep 1	Rep 2	Rep 3	Rep 4	Rep 5								Student's-t	Chebychev
1	0-6 inches	0.25	0.09	0.16	0.11			3	0.25	0.120	0.036	0.197	1.646	0.021	0.18	0.21
2	6 - 12 inches	0.25	0.09	0.1	0.09			3	0.25	0.093	0.006	0.032	0.339	0.003	0.10	0.11
3	12 - 24 inches	0.50	0.12	0.16	0.15			3	0.50	0.143	0.021	0.114	0.795	0.012	0.18	0.20
4																
5																
6																
7																
8																
9																
10																
Combine the Weighted DUs		1.00	--	--	--	--	--	9	1.00	0.125	0.014	0.00084	Medium	0.008	0.142	0.160

Must add to 1.00

Degrees of freedom by Welch-Satterthwaite approximation: **4.01**

Chebychev 95% UCL is recommended because the dispersion of the data is high.

Recommended UCL — Chebyshev 95% UCL: **0.160**

The User should consult the instructions for additional guidance on which 95% UCL is recommended for specific data sets.

For calculation purposes only

W^2	SD^2	1/n	W^2*SD^2/n
0.063	0.00130	0.333333333	3.66753E-10
0.063	0.00003	0.333333333	2.41127E-13
0.250	0.00043	0.333333333	6.52006E-10

Notes

DU	Decision unit
SD	Standard deviation
SE	Standard errror
UCL	Upper confidence limit

When there are multiple samples in each strata, the overall mean of the larger DU can be estimated using the following formulae. Let n_i represent the number of samples from region i, $x\text{-}bar_i$ represent the mean of the ISM samples from region i, s_i represent the SD of the replicate ISM samples from region i, and w_i represent the weight, i.e., the relative size associated with region i. Note that if all strata are of the same size, the w_i are equal, and these equations simplify to the more common calculation methods for the mean and standard deviation. The relative size is the percentage of the larger DU that is made up of region i. The weighted mean is thus:

$$\text{Weighted Mean} = \sum_i w_i \bar{x}_i$$

The standard error associated with the weighted mean is:

$$\text{Standard Error} = \sqrt{\sum_i w_i^2 \frac{s_i^2}{n_i}}$$

which has degrees of freedom approximated by the Welch-Satterthwaite approximation (Cochran, 1977):

$$df \approx \frac{\left(\sum_i \frac{w_i^2}{n_i} s_i^2\right)^2}{\sum_i \frac{\left(\frac{w_i^2}{n_i} s_i^2\right)^2}{n_i - 1}}$$